How to select appropriate encoding strategies for producing LOD-Enabled Bibliographic Data

LODE-BD

Marcia L.Zeng
Kent State University
USA

Imma Subirats
FAO of UN

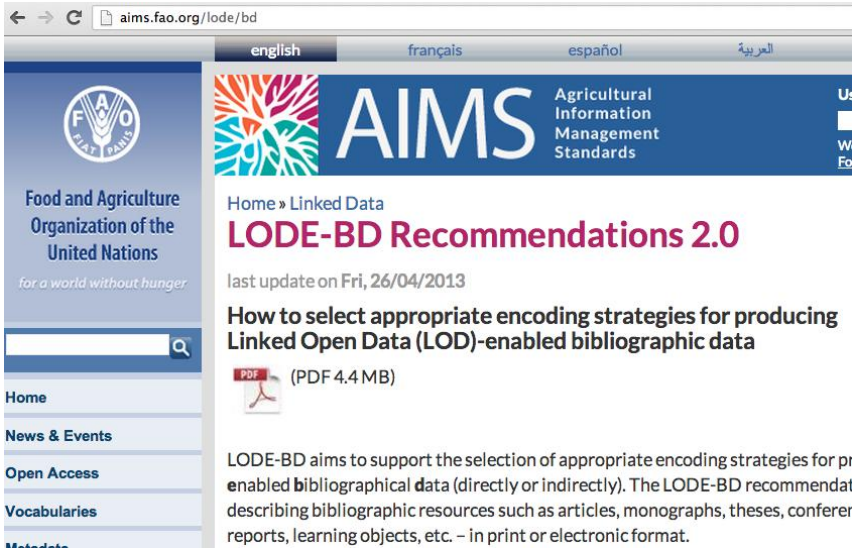
Webinar@AIMS, July 31, 2013

Background

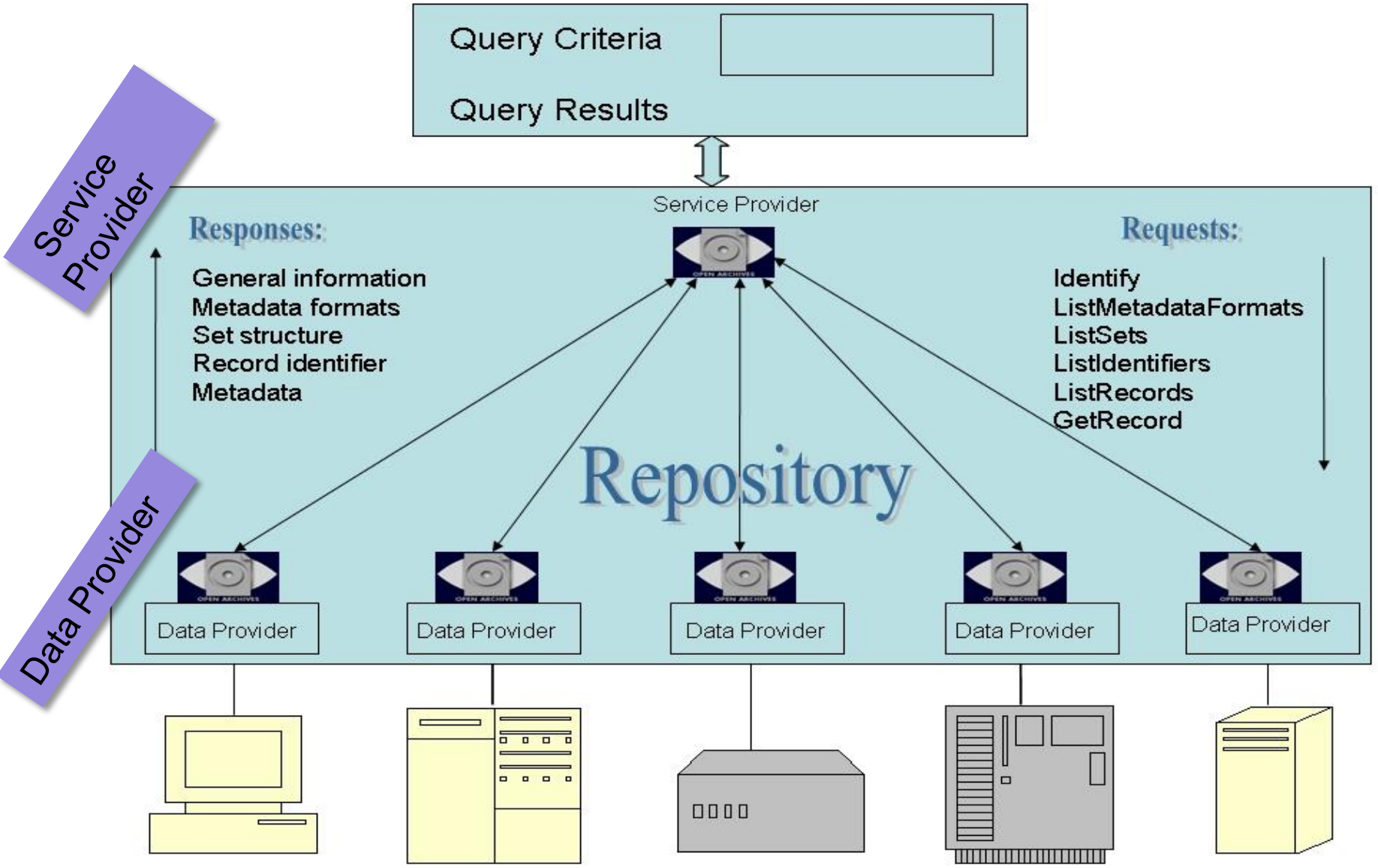
8 repositories
(13 countries, 17 institutions)

Virtual Open Access Agriculture & Aquaculture Repository

Task: turning the bibliographic data
-- from an ad-hoc modeled database in a silo
→ to the data in a standardized metadata repository …
→ heading to the LOD universe



http://aims.fao.org/lode/bd



8 repositories
(13 countries, 17 institutions)

Questions from the data providers:

To make local data into a Linked Data dataset ---

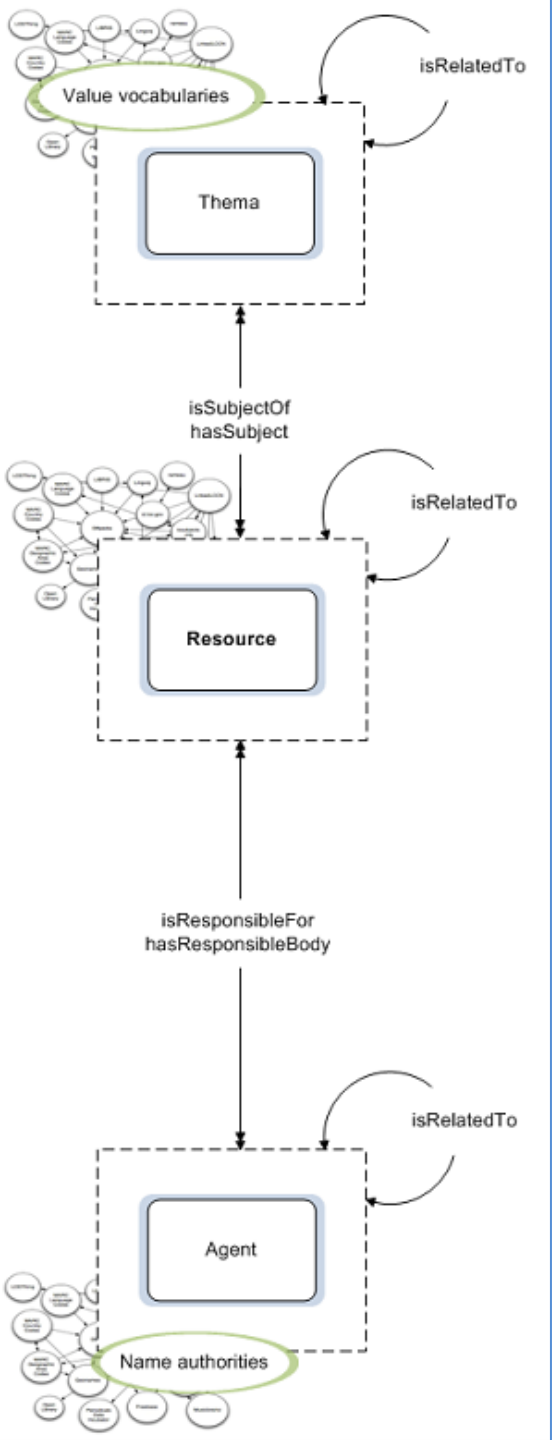
- **What metadata standard(s) to follow?**
- **What is the minimal set of properties to include to insure meaningful data sharing?**
- **Is there any metadata model or application profile for my data?**
- **What data values to provide – plain text string or Unified Resource Identifier (URI)?**
- **How to encode our data in order to move from a local database to a Linked Data dataset?**

Various standards were applied

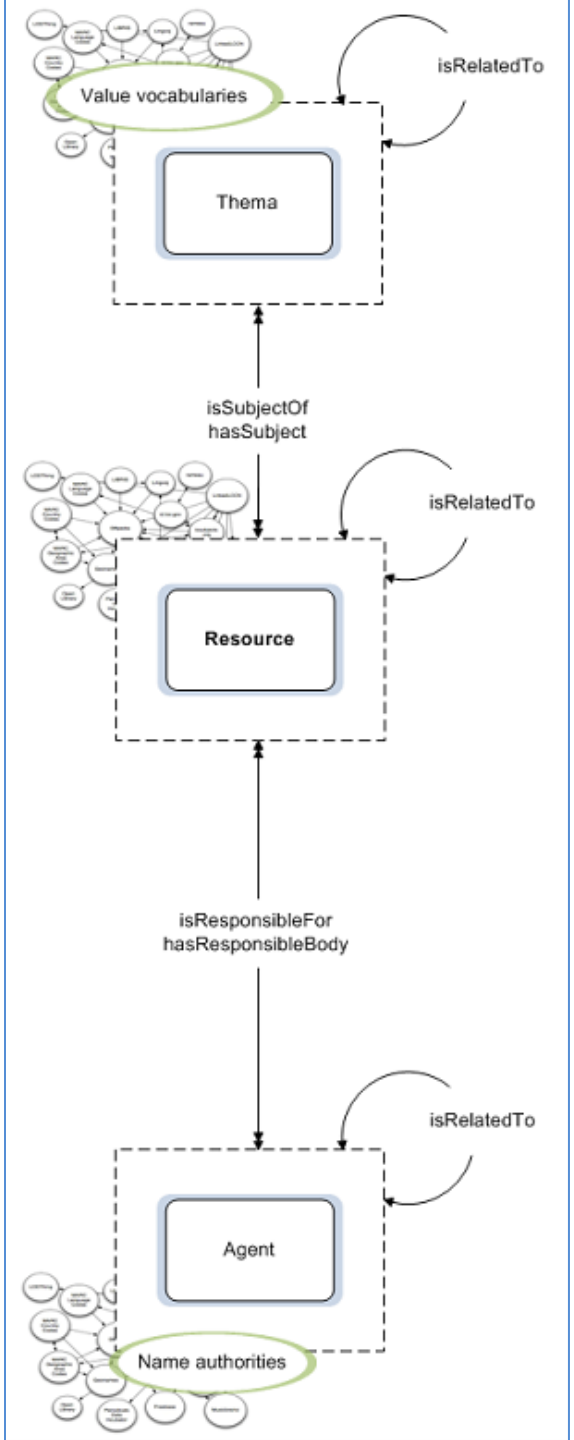
Example: <Responsible Body>- related data fields from 8 data providers

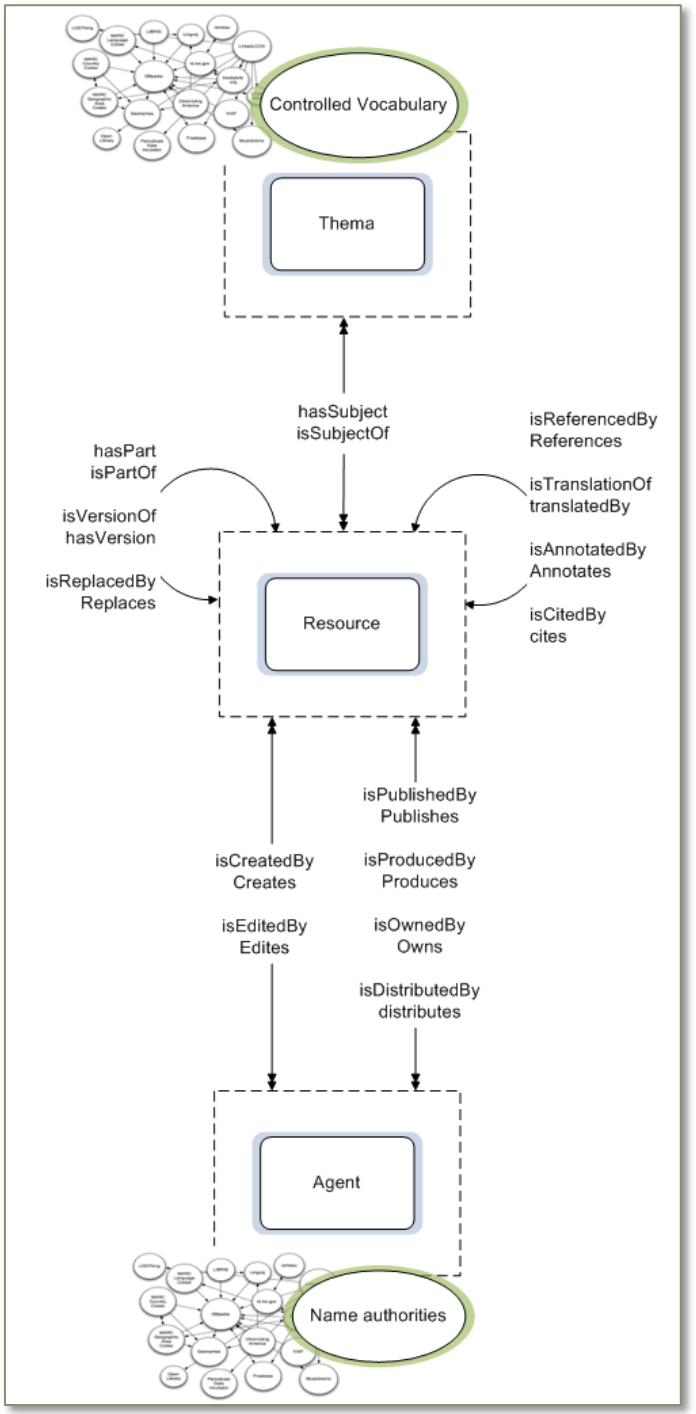
	A	B	C	D	E	F	G	H	I	J
1		1. FAO Open Archive	3. ProdInra - Book/Journal article	4. Organic Eprints	5. Epsilon Open Archive	6. U Gov	7. OceanDocs Repository		8. AGRIS-AgriDrapal e.g.	8 AGRIS DataDictionary
2	creator-person	Personal Name	Personal Name [personal info]	Authors	Creators	Personal Name	Authors	creator-person	Personal Name	Author's name
3			Author Inra							
4			. name							
5		role	.role			role			role	role
6			.[unit detail]							affiliation
7			.[external afficiation detail]							email
8			Author non-inra							
9			. name							
10			.role							
11			. [afficiation detail]			autor list				
12	creator-corporateBody	Corporate Author	Corporate author	Corporate authors			Corporate Authors	creator-corporateBody	Corporate Auth	uthor
13			. name							
14			. type							
15			.[unit detail]							
16			. Acronym							
17			Corporate name							
18	contributor			Editors	Editors		promotors	contributor		
19				research affiliation	Divisions					
20				Funding part	Funders		Funders			
21					Supervisor					
22					Oponent					

- Everyone uses an ad-hoc in-house format.



Step 1. Know your own data model


Value vocabularies
isRelatedTo
Thema
isSubjectOf
hasSubject
isRelatedTo
Resource
isResponsibleFor
hasResponsibleBody
isRelatedTo
Agent
Name authorities


Controlled Vocabulary
Thema
hasSubject
isSubjectOf
isReferencedBy
References
hasPart
isPartOf
isTranslationOf
translatedBy
isVersionOf
hasVersion
isAnnotatedBy
Annotates
isReplacedBy
Replaces
Resource
isCitedBy
cites
isPublishedBy
Publishes
isCreatedBy
Creates
isProducedBy
Produces
isEditedBy
Edites
isOwnedBy
Owns
isDistributedBy
distributes
Agent
Name authorities

Step 2. Map your data fields to the LODE-BD chunks and properties

FAO Open Archive	Controlled
Title	No
Project title	Authority Data
Translated Title	
Transliterated Title	
Subtitle	
Alternative Title	
Personal author	
Corporate author	
Role	
Conference	
Edition	
Publisher	
Place	
Date	
Year	
Copyright date	

A	B	C		D	E
Group	Property	Requirement | M | HR | R | O |		Value Control	Important Attributes
		Non Analytical	Analytical		
1. Title Information	title++	M	M	n	language
	alternative title	O	O	n	
	title supplement	O	O	n	
2. Responsible Body	creator+	HR	HR	n or Name authority (personal, corporate body, conference)	scheme
	contributor	O	O	n or Name authority	
	publisher/issuer+	HR	R	n or Name authority	
3. Physical Characteristics	date++	M	M	Syntax encoding rule	scheme
	identifier+	HR	HR	Syntax encoding rule	scheme
	language++	M	M	Controlled list	scheme
	format/medium+	HR	HR	Controlled list	scheme
	edition /version	R	R	n	
	source+	HR	R	n	

Once

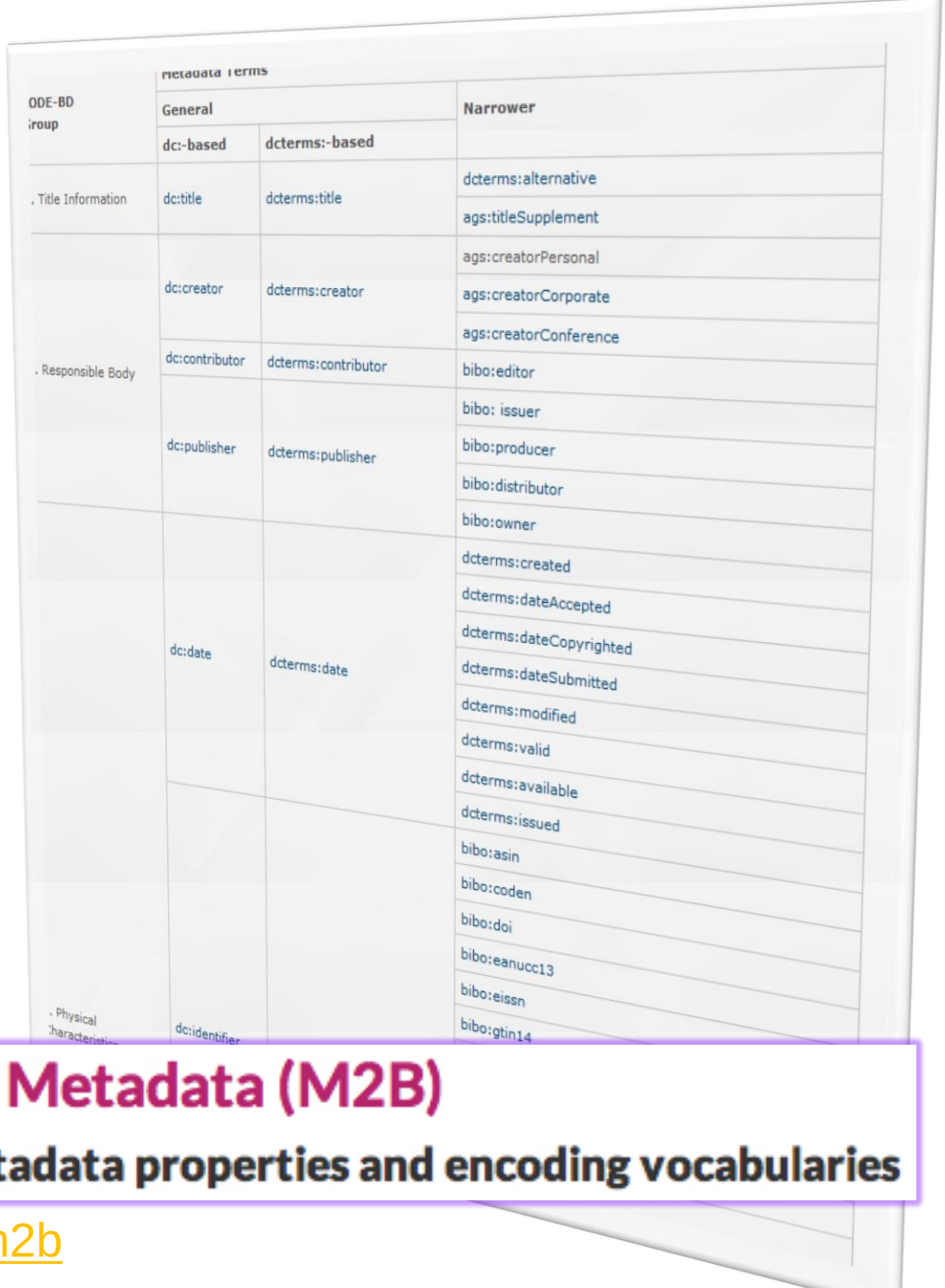
Once or many

0 or once

0 or many

9 Chunks of Properties

1. Title Information
2. Responsible Body
3. Physical Characteristics
4. Location
5. Subject
6. Description of content
7. Intellectual property
8. Usage
9. Relation between documents / agents



ODE-BD iroup	Metadata Terms		
	General		Narrower
	dc:-based	dcterms:-based	
. Title Information	dc:title	dcterms:title	dcterms:alternative
			ags:titleSupplement
	dc:creator	dcterms:creator	ags:creatorPersonal
			ags:creatorCorporate
			ags:creatorConference
. Responsible Body	dc:contributor	dcterms:contributor	bibo:editor
	dc:publisher	dcterms:publisher	bibo: issuer
			bibo:producer
			bibo:distributor
			bibo:owner
	dc:date	dcterms:date	dcterms:created
			dcterms:dateAccepted
			dcterms:dateCopyrighted
			dcterms:dateSubmitted
			dcterms:modified
			dcterms:valid
			dcterms:available
			dcterms:issued
			bibo:asin
			bibo:coden
			bibo:doi
			bibo:eanucc13
			bibo:eissn
. Physical :haracteristi	dc:identifier		bibo:gtin14

Meaningful Bibliographic Metadata (M2B)

Recommendations of a set of metadata properties and encoding vocabularies

http://aims.fao.org/fr/metadata/m2b

"++" = mandatory property
"+" = highly recommended

Minimum properties

Table 1. Groups of Common Properties

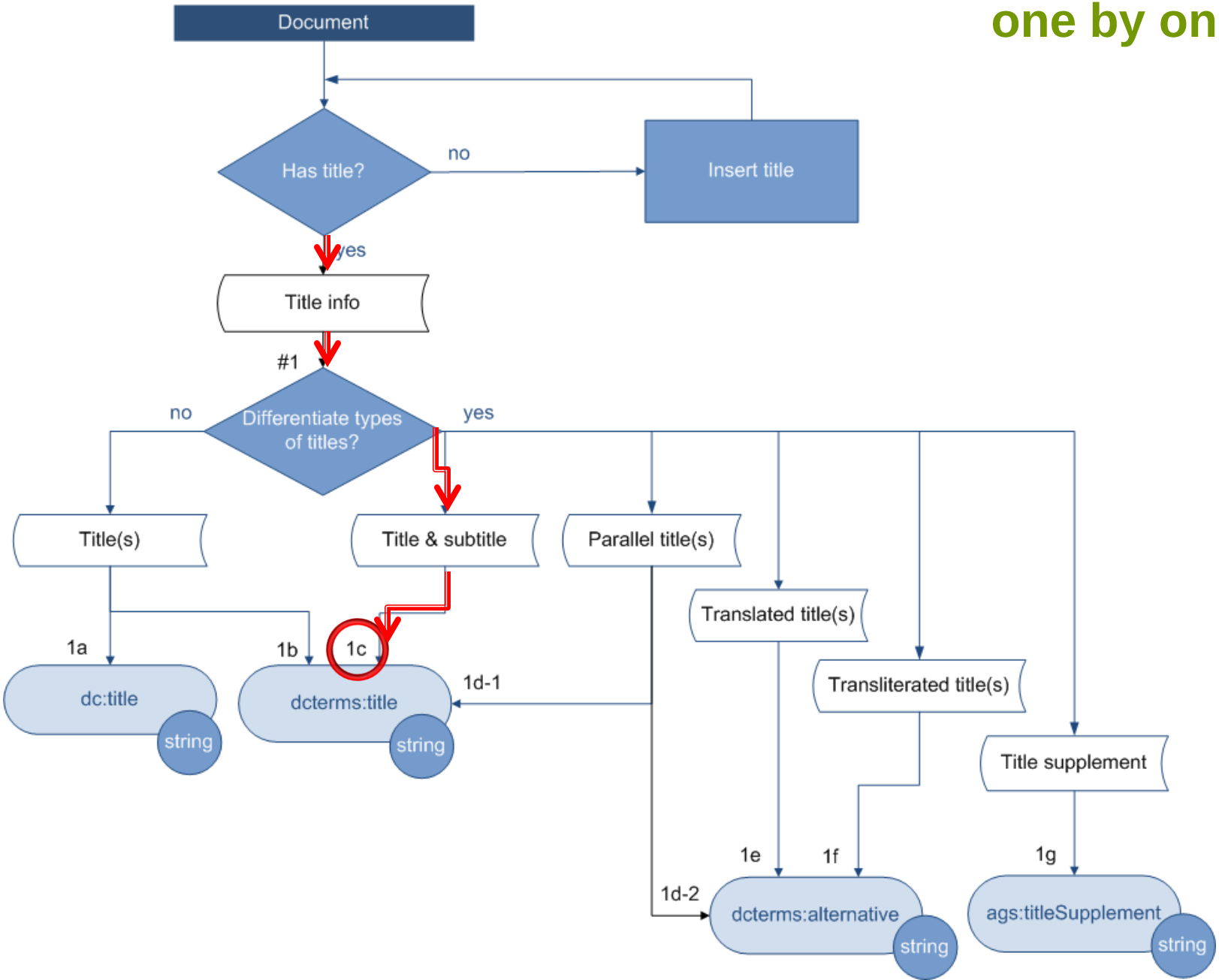
A	B	C		D	E
Group	**Property**	**Requirement | M | HR | R | O |**		**Value Control**	**Important Attributes**
		Non Analytical	**Analytical**		
1. Title Information	**title++**	M	M	n	language
	alternative title	O	O	n	
2. Responsible Body	**creator+**	HR	HR	n or Name authority (personal, corporate body, conference)	scheme
	contributor	O	O	n or Name authority	
	publisher/issuer+	HR	R		

http://aims.fao.org/advice/metadata-beta-version

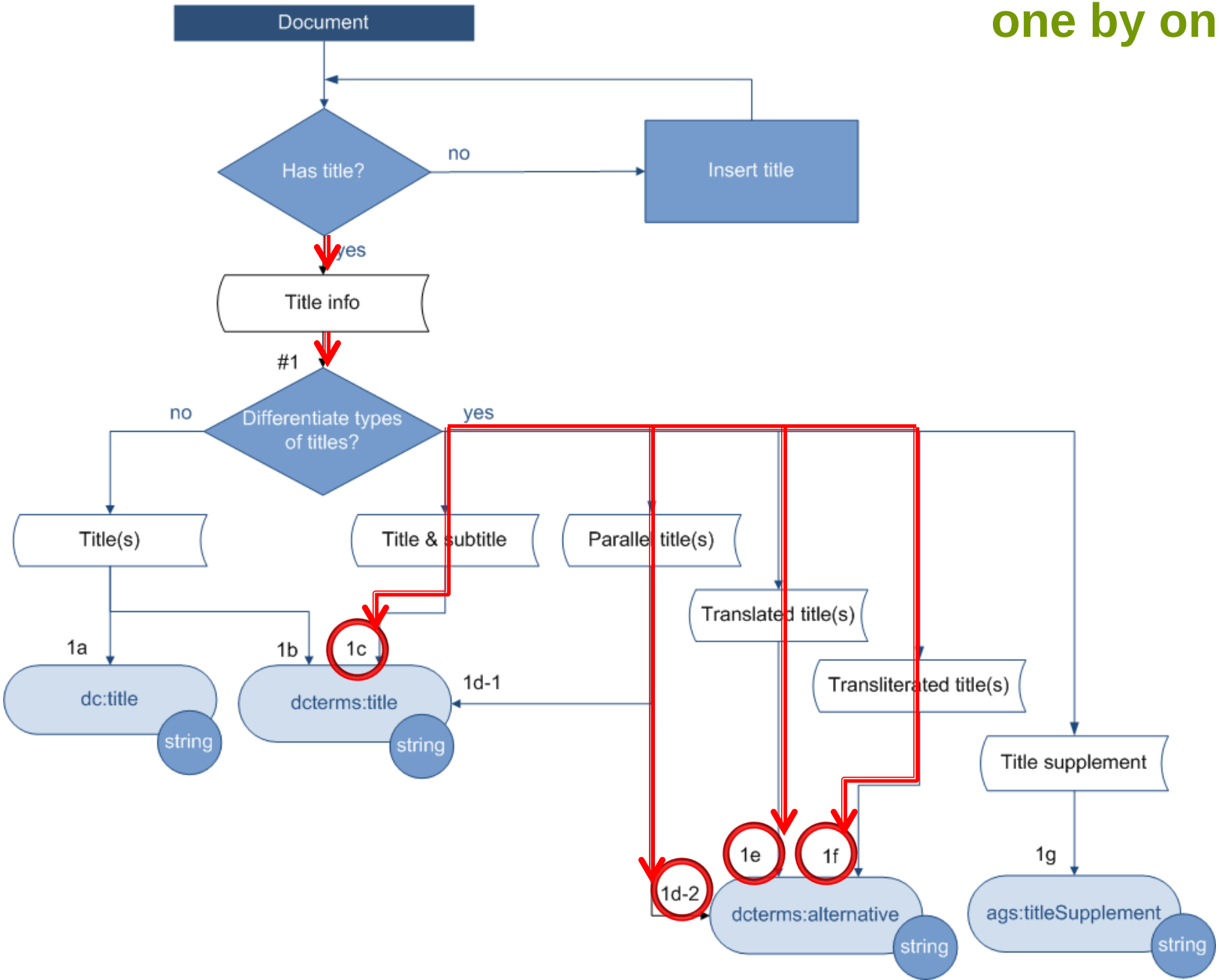
Meaningful Bibliographic Metadata (M2B)

Recommendations of a set of metadata properties and encoding vocabularies

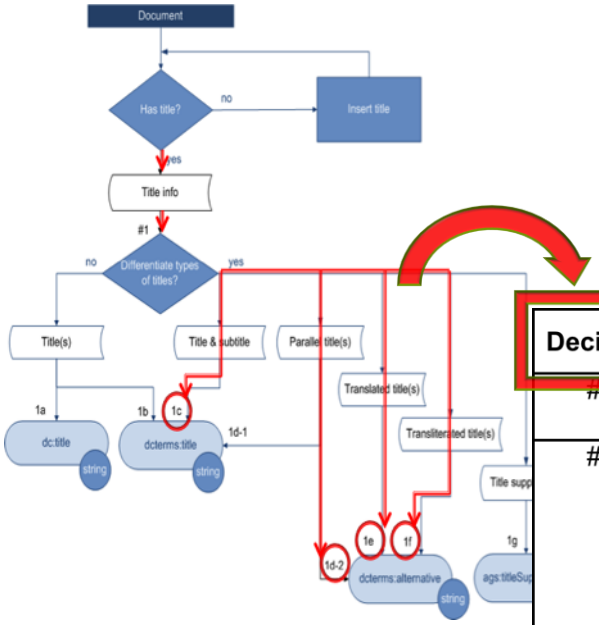
Step 3. Walk through the LODE-BD decision trees, one by one



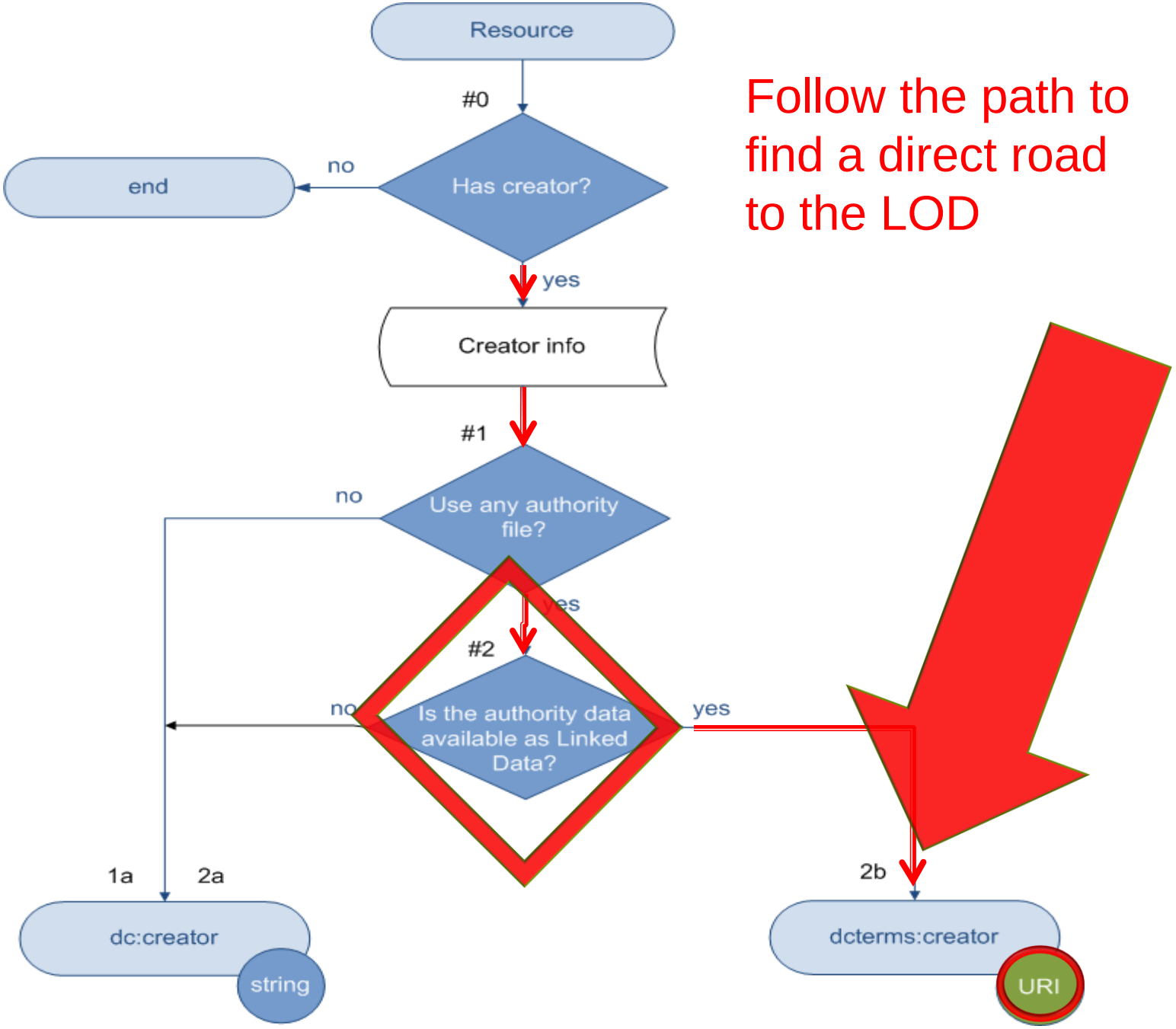
Step 3. Walk through the LODE-BD decision trees, one by one



Step 4. Find metadata terms & the encoding examples



Decision	Question	Answer		Action	Value Type	Examples	
						Metadata Term	Value
#0	Has title?	No		Insert title and go back to #0			
		Yes		Continue to #1			
#1	Differentiate types of titles?	No		1a	String	dc:title	Solar radiation energy and its utilization by Lucerne (Medicagosativa L.) On the state of man [world agricultural situation]
				1b	String	dcterms:title	
		Yes	title(s) and subtitle(s)	1c	String	dcterms:title	FAO yearbook of forest products, 1996-2000
			parallel title(s)	1d-1	String	dcterms:title	Annuaire des produits forestiers de la FAO, 1996-2000
				1d-2	String	dcterms:alternative	
			translated title(s)	1e	String	dcterms:alternative	Anuario de productos forestales de la FAO, 1996-2000 Working together for an International Alliance Against Hunger
			transliterated title(s)	1f	String	dcterms:alternative	Posly dobroj voli Prodovol'stvennoj i Sel'skokhozyajstvennoj Organizatsii Ob'edinennykh Natsij



Follow the path to find a direct road to the LOD

DC Terms User Guide/Publishing Metadata

2.1.11 dcterms:modified

2.1.12 dcterms:title

2.1.13 dcterms:valid

2.2 Properties of the terms namespace used only with non-literal values

2.2.1 dcterms:accrualMethod

2.2.2 dcterms:accrualPeriodicity

2.2.3 dcterms:accrualPolicy

2.2.4 dcterms:accessRights

2.2.5 dcterms:audience

2.2.6 dcterms:conformsTo

2.2.7 dcterms:contributor

2.2.8 dcterms:coverage

2.2.9 dcterms:creator

2.2.10 dcterms:educationLevel

2.2.11 dcterms:extent

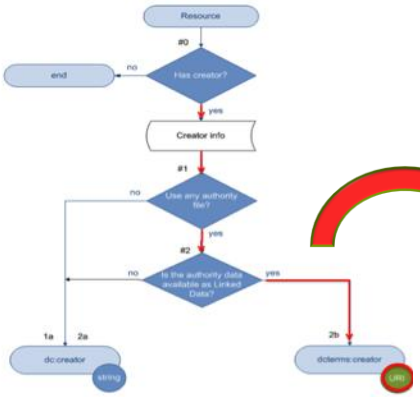
2.2.12 dcterms:format

2.2.13 dcterms:hasFormat

2.2.14 dcterms:hasPart

These properties can only use non-literal values

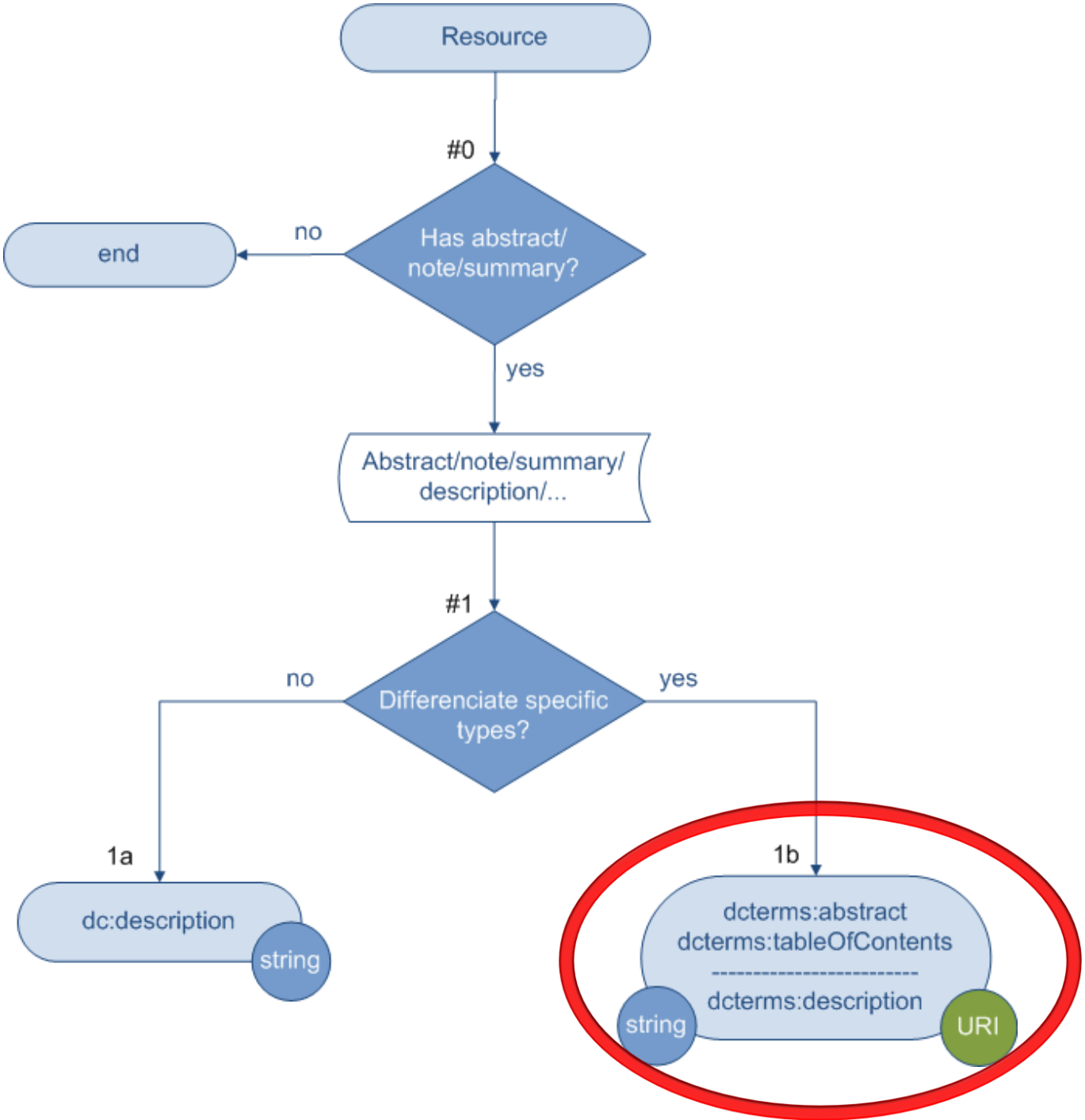
http://wiki.dublincore.org/index.php/User_Guide/Publishing_Metadata



See next slide for a larger image

Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Has creator?	No	End			
		Yes	Continue to #1			
#1	Use any authority file?	No	1a	String	dc:creator	[Unauthorized form]: • Tim Berners-Lee • Tim B-L • Timothy John Berners-Lee • FAO of the UN • FAO Council (78th Session) Nov. 24, 1980, Rome, Italy
		Yes	Go to #2			
#2	Is the authority data available as Linked Data?	No	1b	String	dc:creator	[Authorized form]: • Berners-Lee, Tim • Food and Agriculture Organization of the United Nations • FAO Council (Sess. 78 : 24 Nov 1980 : Rome, Italy)
		Yes	2b	URI	dcterms:creator	• http://aims.fao.org/aos/corporate/c_1297 [1] • http://aims.fao.org/aos/conference/c_1842[2] • http://viaf.org/viaf/85312226/#Berners-Lee,_Tim [3]

Decision	Question	Answer	Action	Value Type	Examples	
					Metadata Term	Value
#0	Has creator?	No	End			
		Yes	Continue to #1			
#1	Use any authority file?	No	1a	String	dc:creator	[Unauthorized form]: • Tim Berners-Lee • Tim B-L • Timothy John Berners-Lee • FAO of the UN • FAO Council (78th Session) Nov. 24, 1980, Rome, Italy
		Yes	Go to #2			
#2	Is the authority data available as Linked Data?	No	1b	String	dc:creator	[Authorized form]: • Berners-Lee, Tim • Food and Agriculture Organization of the United Nations • FAO Council (Sess. 78 : 24 Nov 1980 : Rome, Italy)
		Yes	2b	URI	dcterms:creator	http://aims.fao.org/aos/corporate/c_1297 [1] • http://aims.fao.org/aos/conference/c_1842[2] • http://viaf.org/viaf/85312226/#Berners-Lee,_Tim [3]


Resource
#0
Has abstract/ note/summary?
no
end
yes
Abstract/note/summary/ description/...
#1
Differenciate specific types?
no
yes
1a
dc:description
string
1b
dcterms:abstract
dcterms:tableOfContents
dcterms:description
string
URI

A list of all metadata terms used in LODE-BD and a crosswalk to Schema.org terms

LODE-BD Group	Metadata Terms			schema.org Terms	schema.org Types
	General Metadata Terms		More Specific Terms dcterms: + bibo:	> : narrowMatch < : boardMatch	[Dots indicate the level of a sub-type]
	dc:-based	dcterms:-based			
1. Title Information	dc:title	dcterms:title	dcterms:alternative	name	Thing
				> headline	. CreativeWork
				> alternativeHeadline	. CreativeWork
2. Responsible Body	dc:creator	dcterms:creator		creator author	. CreativeWork
	dc:contributor	dcterms:contributor	bibo:editor	contributor > editor	. CreativeWork
				> Illustrator	. . Book
	dc:publisher	dcterms:publisher	bibo:issuer	publisher	. CreativeWork
			bibo:producer	< provider	
			bibo:distributor		. CreativeWork
			bibo:owner		
3. Physical Characteristics	dc:date	dcterms:date	dcterms:created	dateCreated	. CreativeWork
			dcterms:dateAccepted		
			dcterms:dateCopyrighted	copyrightYear	. CreativeWork
			dcterms:dateSubmitted		
			dcterms:modified	dateModified	. CreativeWork
			dcterms:valid		

http://aims.fao.org/lode/bd-2/schema-org-crosswalk

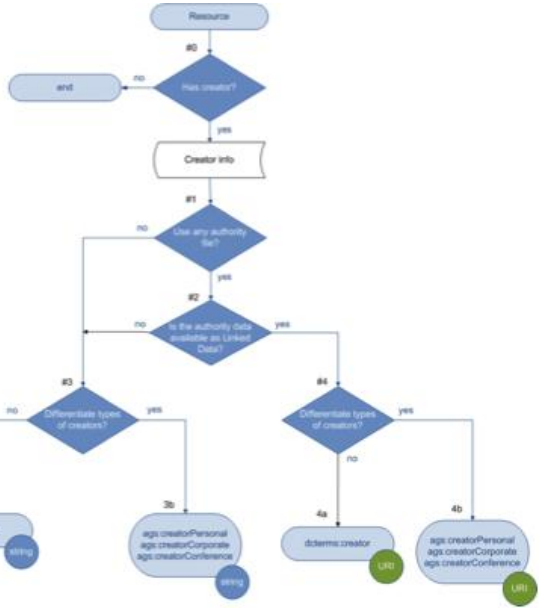
Summary：LODE－BD Method

Various Data Dictionaries & Sample Records

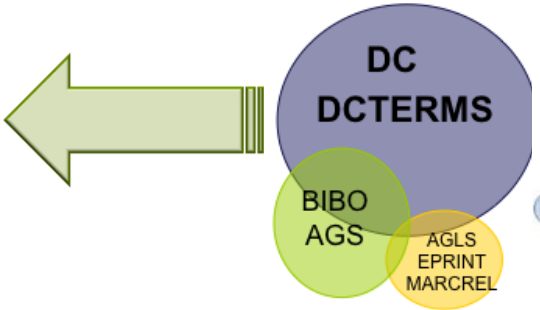
9 Chunks of Properties

Group	Property	Requirement | M | HR | R | O | Non Analytical	Requirement | M | HR | R | O | Analytical	Value Control	Important Attributes
1. Title Information	title++	M	M	n	language
	alternative title	O	O	n	
	title supplement	O	O	n	
2. Responsible Body	creator+	HR	HR	n or Name authority (personal, corporate body, conference)	scheme
	contributor	O	O	n or Name authority	
	publisher/issuer+	HR	R	n or Name authority	
3. Physical Characteristics	date++	M	M	Syntax encoding rule	scheme
	identifier+	HR	HR	Syntax encoding rule	scheme
	language++	M	M	Controlled list	scheme
	format/medium+	HR	HR	Controlled list	scheme
	edition /version	R	R	n	
4. Location	source+	HR	R	n	
	location++	M	M	n or Rule [Holding unit names may be managed through a controlled list]	
	availability	O	O	n	
5. Subject	subject term+	HR	HR	Controlled vocabulary	language scheme
	classification	O	O	Controlled vocabulary, Classification system	scheme
	[freely assigned] keyword	R	R	n	language
	geographic term	O	O	Controlled vocabulary	language scheme
6. Description of content	description/abstract (or note/ summary/ table of contents)	R	R	n	language
	type/form/genre	R	R	Controlled vocabulary	language scheme
7. Intellectual property	rights+ term of use access condition	R	R	n [Rights holders may be managed through name authorities]	
8. Usage	audience	O	O	Controlled list	scheme
	literary indication	O	O	Controlled list	scheme
	education level	O	O	Controlled list	scheme
9. Relation	[relation between resources]+	O	HR	Controlled resource IDs	
	[relation between agents]	O	O	n or Name authority	

18 Decision Trees & Scenarios



LODE-BD

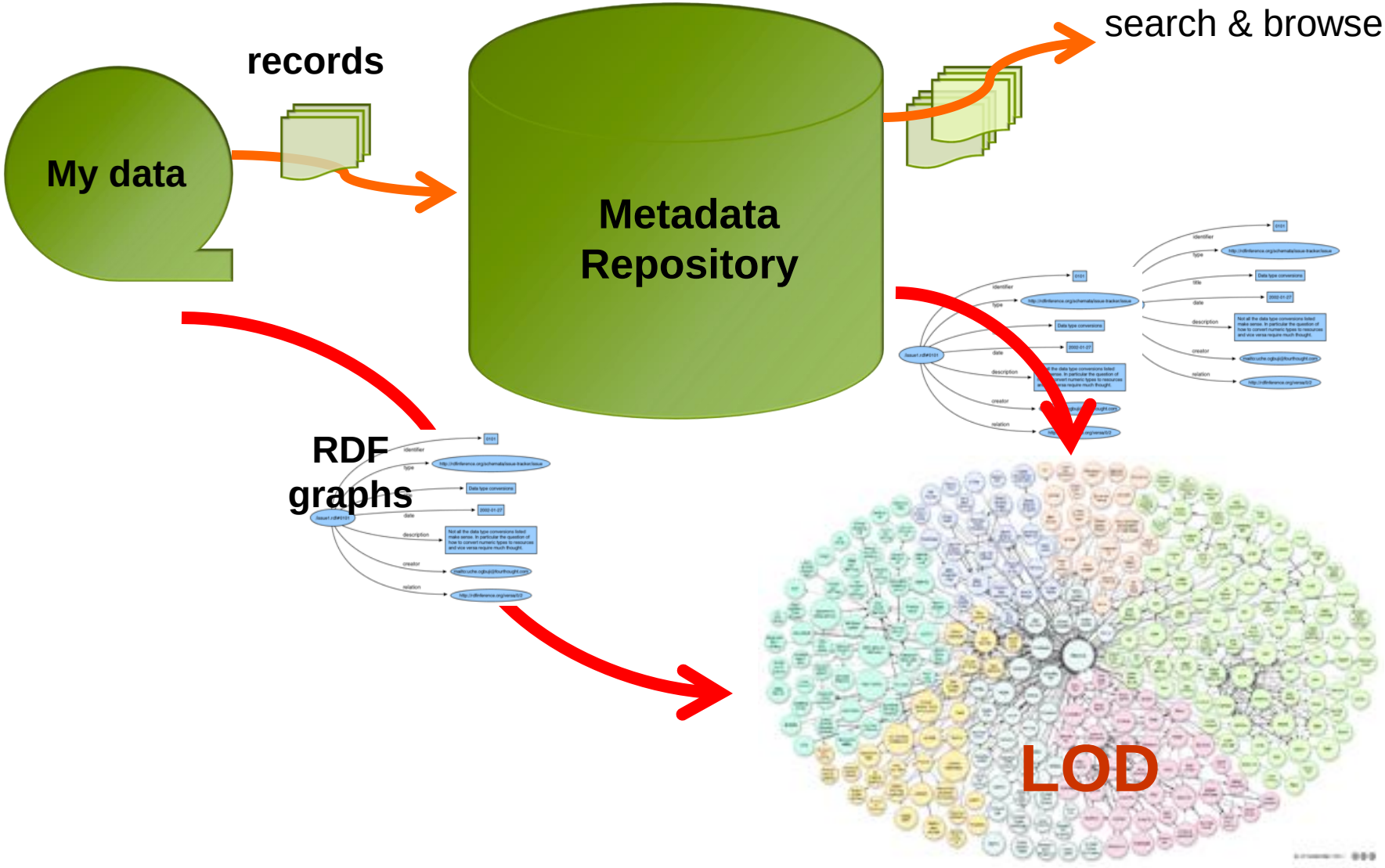


Step 5. Implement the LODE-BD approach in a system

- Alternative #1, "**Design-time**" strategy: go back and change your current ad-hoc model to use the LODE “good practices” model. = > This means some changes to your database and the services that access it.

- Alternative #2, "**Run-time**" strategy: you convert on the fly to a “good-practices” model upon request and leave your ad-hoc model unchanged. => This means adding a conversion service.

Step 6. Output your data



LODE-BD principles

- To promote the use of **well-established metadata standards** and the emerging LOD-enabled vocabularies proposed in the Linked Data community;
- To encourage the use of **authority data, controlled vocabularies**, and **syntax encoding standards** in metadata statements whenever possible;
- To encourage the use of **resource URIs** as data values when they are available;
- To facilitate the decision-making process regarding data encoding for the purpose of **exchange and reuse**;
- To provide a reference support that is **open** for suggestions of new properties and metadata terms according to the needs of the Linked Data community.

THANK YOU!

- LODE-BD Recommendations v.2.0
- http://aims.fao.org/lode/bd

See also:
AGROVOC as LOD, http://aims.fao.org/standards/agrovoc/about
OpenAgris bibliographic data as LOD. http://aims.fao.org/openagris